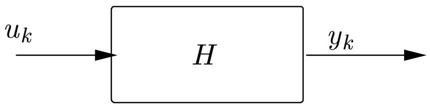
Chapter 6

FIR Digital Filter Design

A digital filter is simply a system that transforms input signals to output signals, generally with different frequency characteristics. In the following figure



u_k is the input, y_k is the output and H is the filter, a linear time-invariant discrete time system with impulse response coefficients h_k. In contrast to the situation in control systems, here the transfer function H is free to be chosen. From the frequency domain point of view, we note that for $u_k = e^{i\omega k}$,

$$y_k = \sum_{l=-\infty}^{\infty} h_l e^{i\omega(k-l)} = H(e^{i\omega})u_k$$

so that the output frequency characteristics can be directly modified by $H(e^{i\omega})$. Designing a transfer function to produce output with desired frequency characteristics is referred to as digital filter design.

Two general classes of filters of interest are:

Finite impulse response (FIR) or nonrecursive filters (often noncausal)

$$H(e^{i\omega}) = \sum_{l=-N}^{N} h_l e^{-i\omega l}$$

This corresponds to the input-output equation

$$y_k = \sum_{l=-N}^{N} h_l u_{k-l}$$

Infinite impulse response (IIR) or recursive filters (almost always causal)

$$H(z) = \frac{b_0 + b_1 z^{-1} + \ldots + b_m z^{-m}}{1 + a_1 z^{-1} + \ldots + a_n z^{-n}}$$

which is equivalent to the difference equation

$$y_k + a_1 y_{k-1} + \cdots + a_n y_{k-n} = b_0 u_k + \cdots + b_m u_{k-m}$$

A common example of an FIR filter is:

Averaging using $2N + 1$ points

$$y_k = \frac{1}{2N+1} \sum_{l=-N}^{N} u_{k+l}$$

corresponding to the filter

$$H(\omega) = \frac{1}{2N+1}\frac{\sin(N+1/2)\omega}{\sin\omega/2}$$

Larger N generates more zeros $\Rightarrow$ more rapid the oscillation

We shall discuss one approach to the design of FIR filters which makes good use of Fourier analysis.

The Fourier Approach to FIR Digital Filter Design

Expand the desired frequency response $H_d(e^{i\omega})$ into its Fourier series representation:

$$H(e^{i\omega}) = \sum_{j=-\infty}^{\infty} h_j^d e^{-ij\omega}$$

Choose filter coefficients as h_j^d. Can in principle give any frequency response.

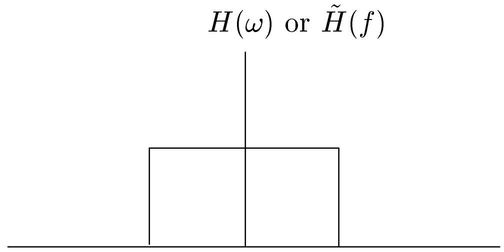
However, one must truncate to give a usable filter

$$H(e^{i\omega}) = \sum_{j=-N}^{N} h_j^d e^{-ij\omega}$$

This approach has the following property of

Optimality: From Fourier series, for a filter with $2N+1$ coefficients, $H(e^{i\omega})$ is s.t. $\frac{1}{2\pi}\int_{-\pi}^{\pi} |H_d(e^{i\omega}) - H(e^{i\omega})|^2 d\omega$ is min.

Apply this to the design of low pass filters



(Here, when the frequency response is real, we often write is as $H(\omega)$, rather than $H(e^{i\omega})$. Also we use $\tilde{H}(f) = H(2\pi f)$).

Becuase of the sharp cutoff in $H(e^{i\omega})$, we know from Fourier series analysis that Gibbs phenomenon gives ripples near discontinuities which are undesirable.

Explore further:

Truncation is the same as multiplication by a sequence w_k with

$$\begin{aligned} w_k &= 1 \quad -N \le k \le N \\ &= 0 \quad |k| > N. \end{aligned}$$

New sequence is $h_k = w_k h_k^d \Rightarrow$

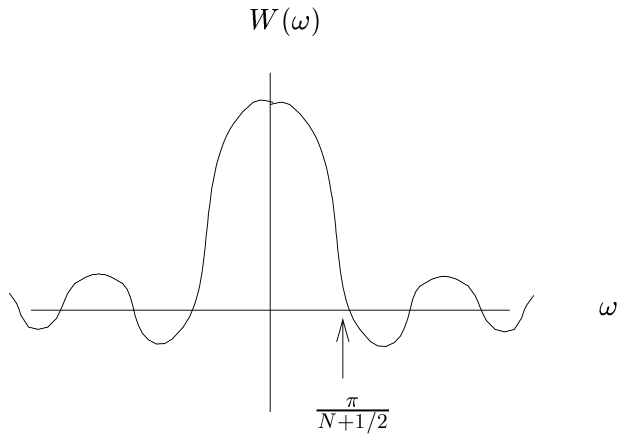
$$H(\omega) = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(\gamma) W(\omega - \gamma) d\gamma$$

Frequency response of

$$W(\omega) \quad = \quad \sum_{k=-N}^{N} e^{-ik\omega}$$

$$
\begin{aligned}
&= e^{-iN\omega} + \ldots + e^{iN\omega} \\
&= e^{-iN\omega}(1 + \ldots + e^{i2N\omega}) = e^{-iN\omega}\frac{1 - e^{+i(2N+1)\omega}}{1 - e^{i\omega}} \\
&= \frac{e^{i(N+1)\omega} - e^{-iN\omega}}{e^{i\omega} - 1} = \frac{e^{i(N+1/2)\omega} - e^{-i(N+1/2)\omega}}{e^{i\frac{\omega}{2}} - e^{-i\frac{\omega}{2}}} \\
&= \frac{\sin(N+1/2)\omega}{\sin\omega/2} \longleftarrow 2\pi \text{ times the Dirichlet kernel!}
\end{aligned}
$$

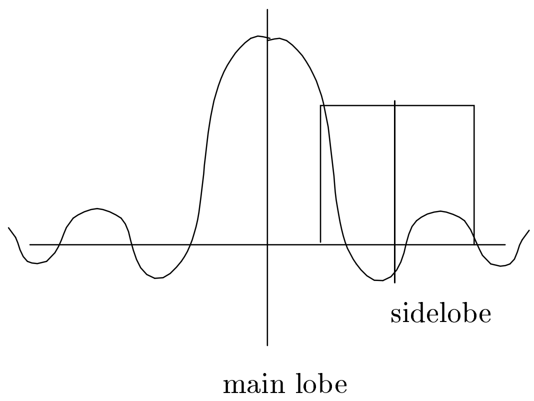
$$W(0) = 2N + 1$$



Either w_k or $W(\omega)$ is called a window. The window $w_k = 1$, $|k| \leq N$, =0 otherwise, is called the rectangular window. We can now interpret Gibbs phenomenon from Fourier series as follows:

$G(\omega)$ is the convolution of the window with the ideal filter frequency response.

Graphically,



If

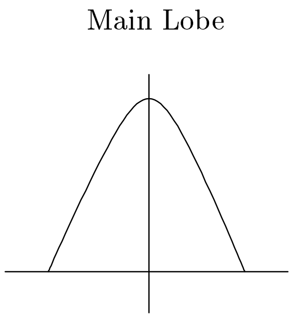
$$W(\omega) = 2\pi\delta(\omega) \quad (w_k = 1 \text{ all } k)$$

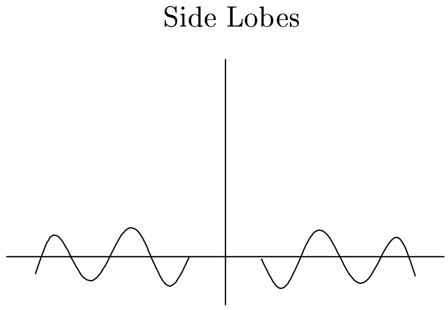
then

$$H(\omega) = H_d(\omega)$$

Since the rectangular window is not perfectly sharp, ripples result.

Split into



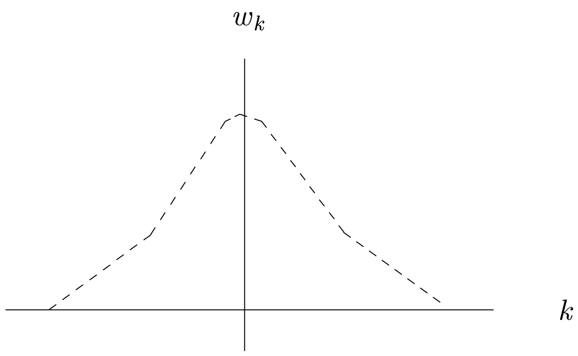


We can give the following interpretation as low pass filter passes through in graphical convolution: main lobe gives transition, side lobes give ripples.

Ideally, want narrow high main lobe, small side lobes, but these properties are conflicting. Now width of main lobe can be more easily controlled. Lengthening the filter reduces the width. So amplitude of side lobes is the main concern.

Instead of having a sharp cutoff, taper the window weights w_k to vanish more gradually at $\pm N$. One such window is the von Hann or Hanning window.

$$\begin{aligned} w_k &= \tfrac{1}{2}(1+\cos\tfrac{\pi k}{N}) \quad |k| \le N \\ &= 0 \qquad\qquad\qquad\; |k| > N \end{aligned}$$



The window weights w_k vanish at the two ends $\pm N$, and considering k as a continuous variable, are tangent there.

Formally,

$$\frac{dw_k}{dk} = -\frac{1}{2}\frac{\pi}{N}\sin\frac{\pi k}{N} = 0 \text{ at } k = \pm N$$

Possible to show

$$W(f) = \frac{\sin 2\pi N f \cos \pi f}{2 \sin \pi f} \frac{1}{1 - \left(\frac{\sin \pi f}{\sin \frac{\pi}{2N}}\right)^2}$$

Begin: <u>Optional Material for the Technically Minded</u>

$$\begin{aligned} W(f) &= \frac{1}{4}\sum_{k=-N}^{N}(e^{i\pi k/N} + 2 + e^{-i\pi k/N})e^{2\pi i k f} \\ &= \frac{1}{4}\sum_{k=-N}^{N}(e^{i\frac{\pi k}{N}} + 2 + e^{i\frac{\pi k}{N}})(1 - \delta_{k,N}\cdot\frac{1}{2} - \frac{1}{2}\delta_{k,-N})e^{2\pi i k f} \\ &\sum_{k=-N}^{N}[e^{2\pi i k f}(1 - \frac{1}{2}(\delta_{k,N} + \delta_{k,-N}))] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=-N}^{N} e^{2\pi ikf} - \frac{1}{2}(e^{2\pi iNf} + e^{-2\pi ifN}) \\
&= \frac{\sin(N+\frac{1}{2})\omega}{\sin\frac{\omega}{2}} - \cos N\omega = \frac{\sin N\omega\cos\frac{\omega}{2} + \cos N\omega\sin\frac{\omega}{2}}{\sin\frac{\omega}{2}} - \cos N\omega \\
&= \frac{\sin N\omega\cos\frac{\omega}{2}}{\sin\frac{\omega}{2}}
\end{aligned}
$$

$$
\begin{aligned}
\therefore W(f) &= \frac{1}{4}\sum_{k=-N}^{N}[e^{2\pi ik(f+\frac{1}{2N})} + 2e^{2\pi ikf} + e^{2\pi ik(f-\frac{1}{2N})}] \\
&= \frac{1}{4}\left\{\frac{\sin 2N\pi(f+\frac{1}{2N})\cos\pi(f+\frac{1}{2N})}{\sin\pi(f+\frac{1}{2N})} + 2\frac{\sin 2\pi Nf\cos\pi f}{\sin\pi f}\right. \\
&\quad \left. + \frac{\sin 2\pi N(f-\frac{1}{2N})\cos\pi(f-\frac{1}{2N})}{\sin\pi(f-\frac{1}{2N})}\right\}
\end{aligned}
$$

Since

$$
\begin{aligned}
\sin(2N\pi f+\pi) &= -\sin 2N\pi f \\
\sin(2N\pi f-\pi) &= -\sin 2N\pi f
\end{aligned}
$$

L'Hopital's rule

$$
\begin{aligned}
W(\frac{1}{2N}) &= \lim_{f\to\frac{1}{2N}} \frac{\frac{d}{df}(\sin 2\pi Nf\cos\pi f)}{\frac{d}{df}2\sin\pi f\left(1-\frac{\sin^2\pi f}{\sin^2\frac{\pi}{2N}}\right)} \\
&= \lim_{f\to\frac{1}{2N}} \frac{2\pi N\cos 2\pi Nf\cos\pi F}{2\sin\pi f\left(-\frac{1}{\sin^2\frac{\pi}{2N}}2\pi\sin\pi f\cos\pi f\right)} \\
&= \frac{N}{2}
\end{aligned}
$$

we get

$$
W(f) = \frac{\sin 2N\pi f}{4}[-\cot\pi(f+\frac{1}{2N}) + 2\cot\pi f - \cot\pi(f-\frac{1}{2N})]
$$

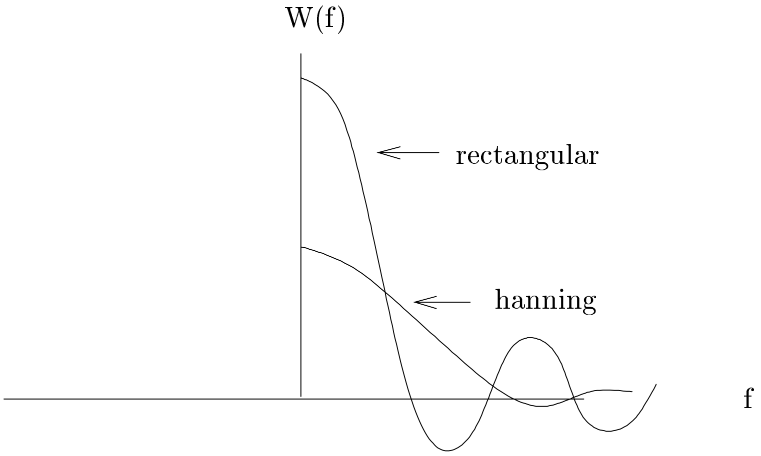
But

$$
\cot(a+b) + \cot(a-b) = \frac{2\sin a\cos a}{\sin^2 a - \sin^2 b}
$$

$$
\begin{aligned}
\Rightarrow W(f) &= \frac{\sin 2\pi Nf}{4}\left[\frac{2\cos\pi f}{\sin\pi f} - \frac{2\sin\pi f\cos\pi f}{\sin^2\pi f - \sin^2\frac{\pi}{2N}}\right] \\
&= \frac{\sin 2\pi Nf\cos\pi f}{2}\left[\frac{-\sin^2\frac{\pi}{2N}}{(\sin\pi f)(\sin^2\pi f - \sin^2\frac{\pi}{2N})}\right] \\
&= \frac{\sin 2\pi Nf\cos\pi f}{2\sin\pi f}\frac{1}{1-\left(\frac{\sin^2\pi f}{\sin^2\frac{\pi}{2N}}\right)}
\end{aligned}
$$

End: <u>Optional Technical Material</u>

The frequency response plots show



Thus at

$$\begin{array}{ll} f = \pm\dfrac{1}{2N} & W(f) = \dfrac{N}{2} \\ f = 0 & W(0) = N \end{array}$$

1st zero occurs at $f = \dfrac{1}{N}$.

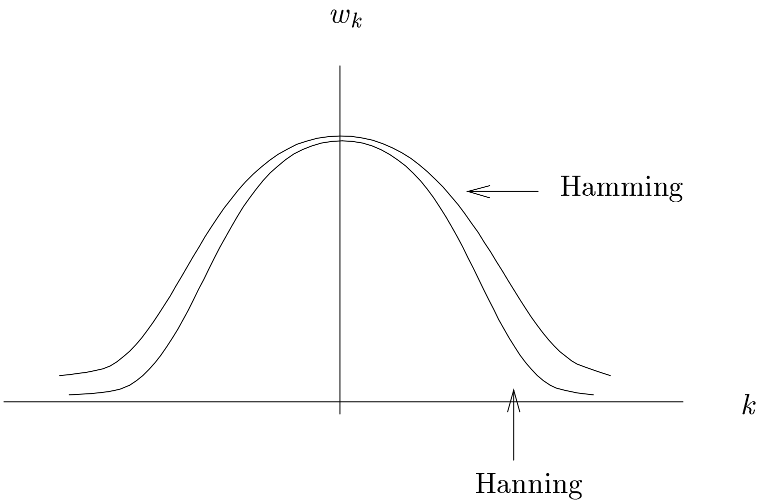
The Hanning window is more than twice as wide as the rectangular window but the side lobe amplitudes are greatly reduced.

Hamming window:

Hamming observed that the Hanning window and the rectangular window have side lobe amplitudes which are opposite in sign. Hence it may be able to reduce the side lobes further by using a combination. Hamming proposed

$$\begin{aligned} w_k &= (2a\tfrac{\cos \pi k}{N} + b) \quad & |k| \leq N \\ &= 0 & \text{otherwise} \end{aligned}$$

with $2a + b = 1$. For $a = 0$, $b = 1$ get rectangular window. $a = \dfrac{1}{4}$ and $b = \dfrac{1}{2}$ gives Hanning window. Hamming applied an optimization routine to choose a and b to reduce the side lobes as much as possible. Turns out a and b are functions of N, but they change very slowly. May take them to be constants. a usually taken to be 0.23, $b = 0.54$.



Let us consider a digital filter design example.

Example: Design of a differentiator:

$$\begin{aligned}
H_d(e^{i\omega}) &= i\omega \qquad -\pi \le \omega \le \pi \\
h_k^d &= \frac{1}{2\pi}\int_{-\pi}^{\pi} i\omega e^{i\omega k} d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} i\omega \frac{de^{i\omega k}}{ik} \\
&= \frac{1}{2\pi k}[\omega e^{i\omega k}]_{-\pi}^{\pi} - \frac{1}{2\pi k}\int_{-\pi}^{\pi} e^{i\omega k} d\omega \\
&= \frac{1}{2\pi k}(\pi e^{ik\pi} + \pi e^{-ik\pi}) - \frac{1}{2\pi k}\left[\frac{e^{i\omega k}}{ik}\right]_{-\pi}^{\pi} \\
&= \frac{\cos k\pi}{k} - \frac{1}{2\pi i k^2}(e^{ik\pi} - e^{-ik\pi}) \\
&= \frac{\cos k\pi}{k} = \frac{(-1)^k}{k} \qquad k \neq 0
\end{aligned}$$

$$\text{For } k = 0, \quad h_k^d = \frac{i}{2\pi}\int_{-\pi}^{\pi} \omega d\omega = 0$$

For $N = 5$ so that we are using an 11-point filter

$$\begin{aligned}
h_0^d &= 0 \\
h_1^d &= -1 \\
h_2^d &= .5 \\
h_3^d &= -\frac{1}{3} = -0.3333 \\
h_4^d &= \frac{1}{4} = 0.25 \\
h_5^d &= -0.2
\end{aligned}$$

Note that by odd symmetry,

$$\begin{aligned}
h_{-k}^d &= \frac{1}{2\pi}\int_{-\pi}^{\pi} i\omega e^{-i\omega k} d\omega \\
&= -\frac{1}{2\pi}\int_{-\pi}^{\pi} i\gamma e^{i\gamma k} d\gamma \\
&= -h_k^d
\end{aligned}$$

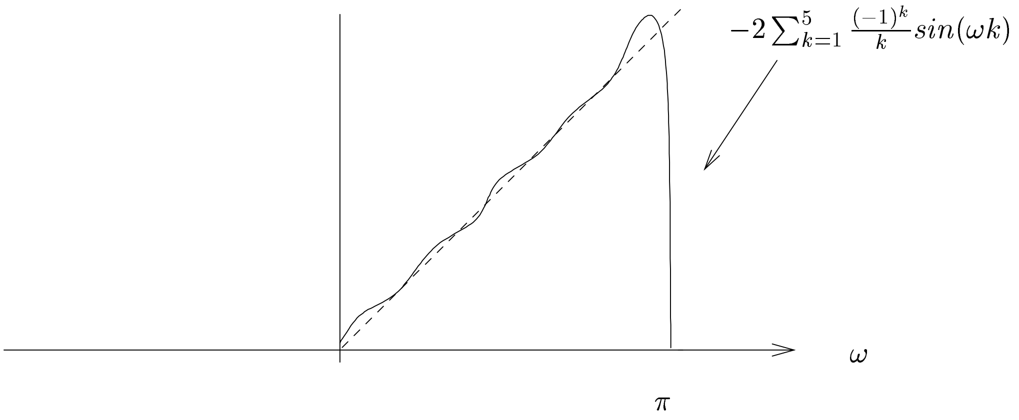
The complete filter is given by:

$$\begin{aligned}
y_n &= 0.2u_{n+5} - 0.25u_{n+4} + 0.3333u_{n+3} - 0.5u_{n+2} + u_{n+1} \\
&\quad - u_{n-1} + 0.5u_{n-2} - 0.3333u_{n-3} + 0.25u_{n-4} - 0.2u_{n-5}
\end{aligned}$$

Frequency response of filter with rectangular window

$$\begin{aligned}
H_r(e^{i\omega}) &= \sum_{k=-5}^{5} \frac{(-1)^k}{k} e^{-i\omega k} \\
&= -2i\sum_{k=1}^{5} \frac{(-1)^k}{k} sin(\omega k)
\end{aligned}$$

At $\omega = \pi$, $\sum_{k=-5}^{5} \frac{1}{k} = 0$ always regardless of how many terms are used! (Why?)



Hamming window:

$$h_k = h_k^d w_k \qquad w_k = 0.54 + 0.46\cos\frac{k\pi}{N}$$

$$\begin{aligned} w_1 &= 0.912148 \\ w_2 &= 0.682148 \\ w_3 &= 0.397852 \\ w_4 &= 0.167852 \\ w_5 &= 0.08 \end{aligned}$$

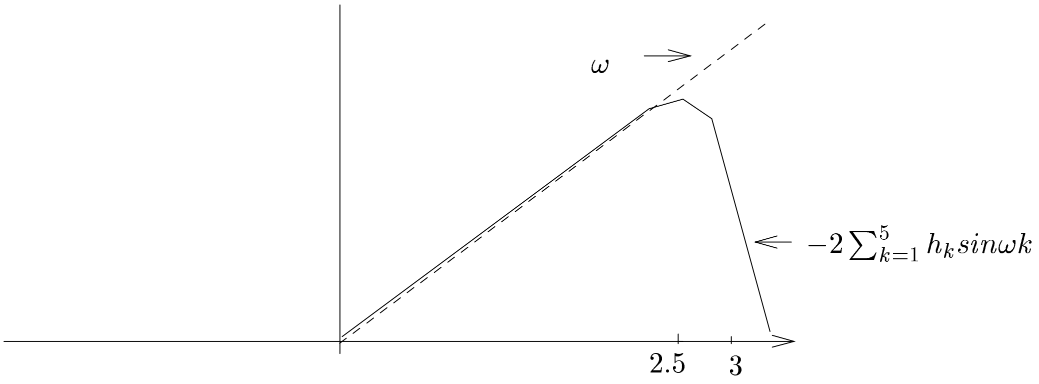
$$h_1 = -0.912148 \quad h_2 = 0.341074 \quad h_3 = -0.132617$$
$$h_4 = 0.041963 \quad h_5 = -0.016$$

The frequency response of the 11-point Hamming-window differentiator is

$$H_h(e^{i\omega}) = \sum \frac{(-1)^k}{k} w_k e^{-i\omega k}$$

At $\omega = \pi$, $\displaystyle\sum_{-N}^{N}\frac{w_k}{k} = 0$. Since w_k is even, h_k will still be odd.

$$\begin{aligned} H_h(e^{i\omega}) &= \sum_{k=-5}^{5} h_k e^{-i\omega k} \\ &= \sum_{1}^{k} h_k e^{-i\omega k} + \sum_{-5}^{-1} h_k e^{-i\omega k} = \sum_{1}^{5} h_k (e^{-i\omega k} - e^{i\omega k}) \\ &= -2i\sum_{1}^{5} h_k \sin\omega k \end{aligned}$$



Kaiser Windows

In Hamming and Hanning windows, there is only one parameter to control the shape of the window, namely, the filter length N. We shall now study a more sophisticated design method, *Kaiser windows*, which allows better control of the ripples.

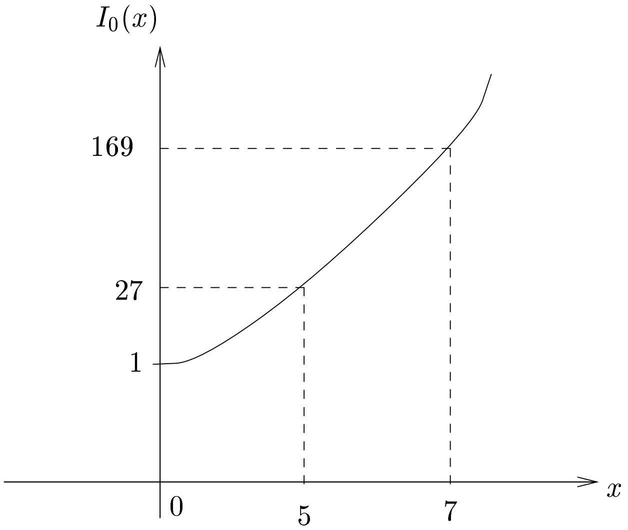
We have seen before that ideally would like w_k and $W(f)$ to be both narrow. The first gives a shorter filter length. The second allows $H(f)$ to approximate $H_d(f)$, the desired filter, more closely. Unfortunately, these are conflicting objectives. Deep Fourier analysis results show that prolate spheroidal functions come close to this ideal. Prompted Kaiser to suggest the weights

$$w_k = \begin{cases} \frac{I_0(\alpha\sqrt{1-\left(\frac{k}{N}\right)^2})}{I_0(\alpha)} & |k| \le N \\ 0 & |k| > N \end{cases}$$

where

$$I_0(x) = 1 + \sum_{n=1}^{\infty} \left[\frac{\left(\frac{x}{2}\right)^n}{n!} \right]^2$$

This is a modified Bessel function of the 1st kind and order zero. (Approximates very well prolate spheroidal functions of order 0.)



$$\begin{aligned} w_k &= \begin{cases} I_0(\alpha\sqrt{1-(k/N)^2})/I_0(\alpha) & |k| \le N \\ 0 & |k| > N \end{cases} \\ w_0 &= 1 \\ w_N &= 1/I_0(\alpha) \end{aligned}$$

$$\alpha - \text{scaling parameter } (1 \lesssim \alpha \lesssim 10)$$

Note that while $I_0(x)$ is an infinite series, it converges rapidly for $n > \frac{x}{2}$, so that it is not necessary to include a large number of terms.

Explanation of why Kaiser window works

Let

$$\begin{aligned} w(t) &= \begin{cases} \frac{I_0[\omega_a\sqrt{\tau^2-t^2}]}{I_0(\omega_a\tau)} & |t| \leq \tau \\ 0 & |t| > \tau \end{cases} \\ &= \begin{cases} \frac{I_0[\omega_a\tau\sqrt{1-(t/\tau)^2}]}{I_0(\omega_a\tau)} & \omega_a\tau = \alpha \\ 0 & \end{cases} \end{aligned}$$

This has Fourier transform

$$\begin{aligned} W(i\omega) &= \frac{2}{I_0(\omega_a\tau)}\frac{\sin[\tau\sqrt{\omega^2-\omega_a^2}]}{\sqrt{\omega^2-\omega_c^2}} \\ &= \frac{2}{I_0(\omega_a\tau)}\frac{\sin(i\tau\omega_a\sqrt{1-(\omega/\omega_a)^2})}{\omega_a i\sqrt{1-(\omega/\omega_a)^2}} \end{aligned}$$

since

$$\sin i\theta = \frac{e^{i(i\theta)}-e^{-i(i\theta)}}{2i} = \frac{e^{-\theta}-e^{\theta}}{2i} = i\sinh\theta$$

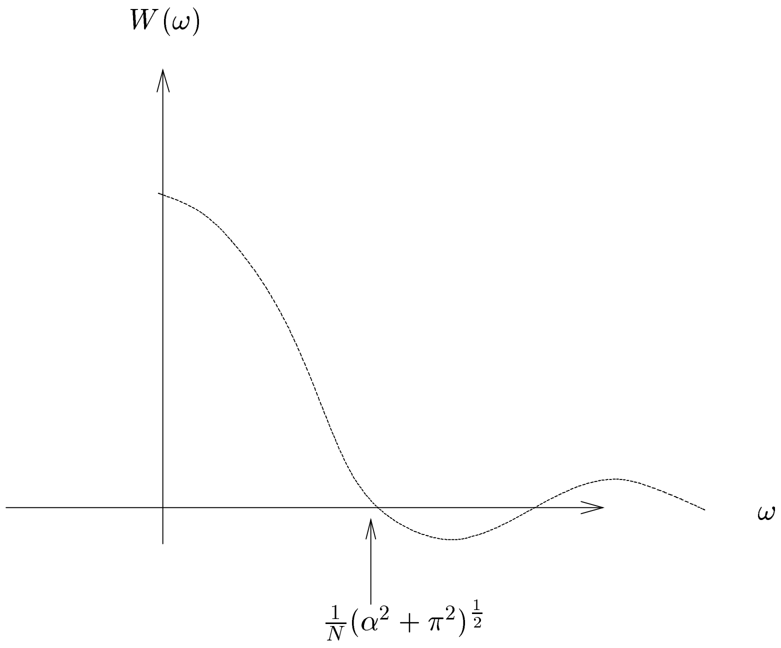
$$\begin{aligned} W(i\omega) &= \frac{2}{I_0(\alpha)}\frac{\sinh[\tau\omega_a\sqrt{1-(\omega/\omega_a)^2}]}{\omega_a\sqrt{1-(\omega/\omega_a)^2}} \\ &= \frac{2\tau}{\alpha I_0(\alpha)}\frac{\sinh[\alpha\sqrt{1-(\omega/\omega_a)^2}]}{\sqrt{1-(\omega/\omega_a)^2}} \end{aligned}$$

Based on the above discussion, we can obtain an approximate frequency response for the Kaiser window by treating the discrete time sequence w_t as a continuous time function $w(t)$. Let $\tau = N$ and $\omega_a = \dfrac{\alpha}{N}$. Taking Fourier transforms gives:

$$W(\omega) = \frac{2N\sinh(\alpha\sqrt{1-(\omega/\omega_a)^2})}{\alpha I_0(\alpha)\sqrt{1-(\omega/\omega_a)^2}}$$

So

$$\begin{aligned} W(0) &= 2N\sinh\alpha/\alpha I_0(\alpha) \\ W(\omega_a) &= 2N/I_0(\alpha) \end{aligned}$$



(This is only an approximation to the actual window function $1 + 2\sum_{1}^{N} w_k \cos(k\omega)$.)

Similarly, for $\omega \geq \omega_a$

$$W(\omega) = \frac{2N \sin[\alpha\sqrt{(\omega/\omega_a)^2 - 1}]}{\alpha I_0(\alpha)\sqrt{(\omega/\omega_a)^2 - 1}}$$

First zero (approx.) at $\alpha\sqrt{(\omega/\omega_a)^2 - 1} = \pi$

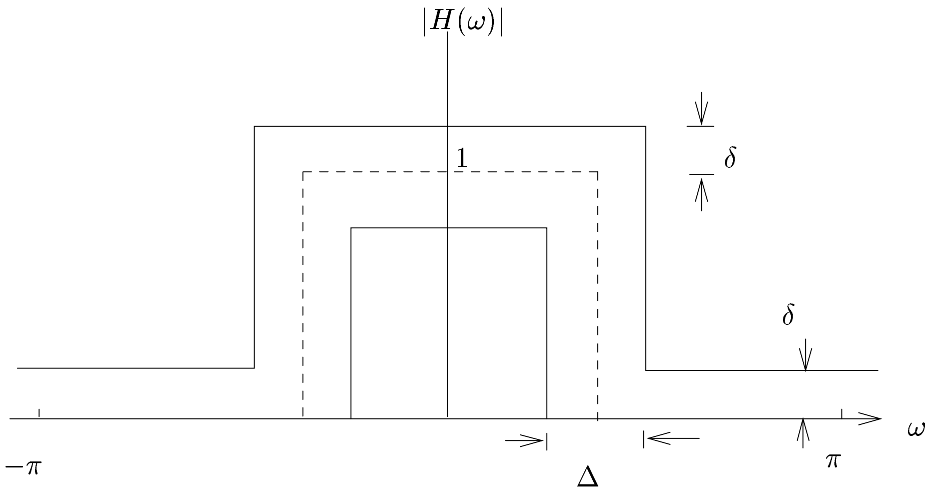
$$\begin{aligned} \alpha^2 \left[\left(\frac{\omega}{\omega_a} \right]^2 - 1 \right] &= \pi^2 \\ \left(\frac{\alpha}{\omega_a} \right)^2 \omega^2 &= \alpha^2 + \pi^2 \\ \Rightarrow \omega &= \frac{1}{N}(\alpha^2 + \pi^2)^{1/2} \end{aligned}$$

If $\alpha = 0$, Kaiser window is identical to the rectangular window, since $w_k = 1, \quad |k| \leq N$.
For $\alpha = \pi\sqrt{3} \doteqdot 5.4414$, first zero at

$$\frac{1}{N}(\alpha^2 + \pi^2)^{1/2} = \frac{1}{N}(4\pi^2)^{1/2} = \frac{2\pi}{N}$$

If we treat similarly the Hamming window weights as a continuous function of t, we get the first zero location for the Hamming window frequency response also at $\omega = \frac{2\pi}{N}$. $\Rightarrow$ Kaiser main lobe width = Hamming main lobe width.. In fact, for $\alpha = \pi\sqrt{3}$, Kaiser and Hamming windows are very close.

To apply the Kaiser window digital filter design procedure systematically, we examine low-pass filter design specifications.



$$\begin{aligned} \Delta &= \text{transition width} \\ \delta &= \text{tolerance on ripple} \end{aligned}$$

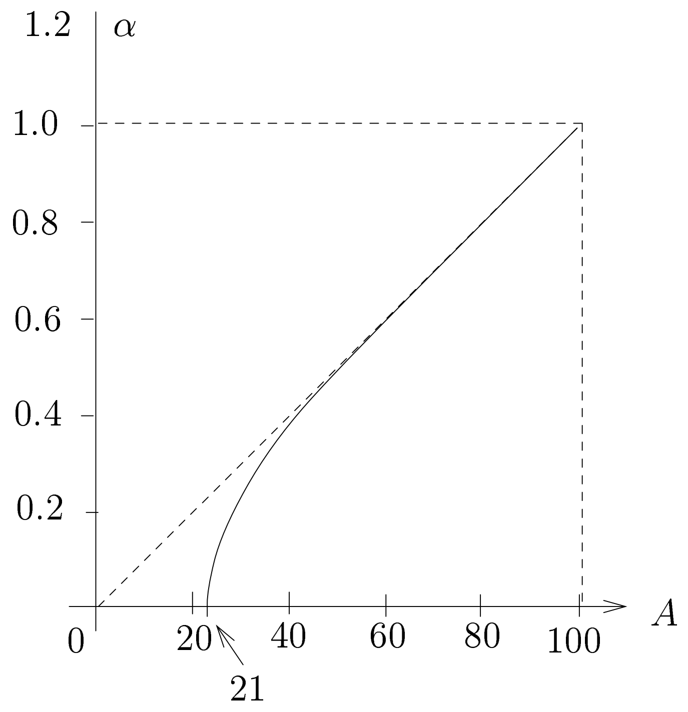
Kaiser Procedure for Determination of the Filter Window

1. $A := -20\log_{10}\delta$ (attenuation in db)
2. $N \geq \dfrac{A - 7.95}{4.57\Delta}$
3. $\alpha = \alpha(A)$ from formula given below:

$$\alpha = \begin{cases} 0.1102(A - 8.7) & A \geq 50 \\ 0.5842(A-21)^{0.4} + 0.07886(A-21) & 21 < A < 50 \\ 0 & A \leq 21 \end{cases}$$

Or look up table:

A	α
25	1.333
30	2.117
35	2.783
40	3.395
45	3.975
50	4.551
55	5.102
60	5.653
65	6.204
70	6.755
75	7.306
80	7.857
85	8.408
90	8.959
95	9.501
100	10.061



4. From N, α, compute $\{w_k\}$.
5. If $\{h_k^d\}$ are ideal low-pass filter coefficients, then

$$h_k = w_k h_k^d.$$

Remark: The parameter A varies slowly with δ: It's much easier (in terms of N) to reduce ripple size than to narrow the transition band.

Example: Suppose we would like to design a low-pass filter with less than .1% ripple and $\Delta = .05\pi$ rad.
Attenuation $A = 60$. Table gives $\alpha = 5.653$.

$$N = \frac{A - 7.95}{4.57 \times 0.05\pi} = 72.51 \rightarrow 73.$$

$\Rightarrow$ need 147 coefficients to meet our design specs.